

ABSTRACT

A document processing system and method for storing and retrieving in an image file a plurality of captured document images for subsequent retrieval on an individual basis. The document processing system includes at least one computer running system software that interfaces with transport hardware to provide document control and capture of document images and document data in various formats. The captured image data is indexed by creating an index file containing indexing data having a structure that is in the form of a self-describing document, wherein elements describe the indexing data for the captured document images to allow subsequent retrieval on an individual basis. In a preferred embodiment, the index file structure is a document type definition (DTD) created using the extensible markup language (XML).